# TARGET DETECTION IMPROVEMENTS USING
# TEMPORAL INTEGRATIONS AND SPATIAL FUSION

## CROSS-REFERENCE

[0001] This application claims the benefit of U.S. provisional application Serial No. 60/456,190, filed March 21, 2003.

## TECHNICAL FIELD

[0002] The present invention relates generally to data fusion, and more particularly to target identification techniques utilizing temporal integrations and spatial fusions of sensor data.

## BACKGROUND OF THE INVENTION

[0003] Sensor systems incorporating a plurality of sensors (multi-sensor systems) are widely used for a variety of military applications including ocean surveillance, air-to-air and surface-to-air defense (e.g., self-guided munitions), battlefield intelligence, surveillance and target detection (classification), and strategic warning and defense. Also, multi-sensor systems are used for a plurality of civilian applications including condition-based maintenance, robotics, automotive safety, remote sensing, weather forecasting, medical diagnoses, and environmental monitoring (e.g., weather forecasting).

[0004] To obtain the full advantage of a multi-sensor system, an efficient data fusion method (or architecture) may be selected to optimally combine the received data from the multiple sensors to generate a decision output. For military applications (especially target

1

recognition), a sensor-level fusion process is widely used wherein data received by each individual sensor is fully processed at each sensor before being output to a system data fusion processor that generates a decision output (e.g., "validated target" or "no desired target encountered") using at least one predetermined multi-sensor algorithm. The data (signal) processing performed at each sensor may include a plurality of processing techniques to obtain desired system outputs (target reporting data) such as feature extraction, and target classification, identification, and tracking. The processing techniques may include time-domain, frequency-domain, multi-image pixel image processing techniques, and/or other techniques to obtain the desired target reporting data.

[0005] It is advantageous to detect or identify image elements or targets as far away as possible. For example, in battle situations, candidate or potential targets should be detected early, increasing the likelihood of an early detection of a target or other object. For a simple background scene such as a blue sky, a target may be recognized from a relatively long range distance. However, for some high clutter situations such as mountains and cities, the detection range is severely reduced. Moreover, such clutter situations are often complicated to process. For example, the background may be mixed with different clutter types and groups. Also the background clutter may be non-stationary. In these types of situations, the traditional constant false alarm ratio (CFAR) detection technique often fails.

[0006] Spatio-temporal fusion for target classification has been discussed in the art. The fusion is conducted in *the likelihood function reading domain*. In general, the likelihood

functions (pdfs) are obtained from training data based on single sensor and single frame measurements. Therefore, fusion is conducted using the likelihood readings of the features extracted from measurements of single sensor and frame, only one set of likelihood functions needs to be stored for a single sensor and frame, no matter how many sensors and frames are used for fusion. On the other hand, if the detection process uses thresholding technique instead of likelihood functions, the features values can be directly fused from different sensors and time frames in *the feature domain* for *target detection.*

[0007] Spatial fusion is defined as the fusion between different sensors, and temporal fusion is defined as the temporal integration across different time frames within a single sensor. Accordingly, it is desirable to develop and compare different spatial fusion and temporal integration (fusion) strategies, including *pre-detection* integration (such as additive, multiplicative, MAX, and MIN fusions), as well as the traditional *post-detection* integration (the persistency test). The pre-detection integration is preferably conducted by fusing the feature values from different time frames before the thresholding process (the detection process), while the post-detection integration is preferably conducted after the thresholding process.

## SUMMARY OF THE INVENTION

[0008] The present invention overcomes the problems described above by using both spatial fusion and temporal integration to enhance target detection (recognition). More specifically, pre-detection temporal integration and spatial fusion techniques are disclosed for enhancing target detection and recognition. These techniques involve different spatio-temporal fusion strategies such as the additive, multiplicative, maximum, and minimum

fusions. In spatial fusion, extracted features from different sensors are fused. In temporal fusion, extracted features across a multiple time frame window are fused and integrated. In addition, a double-thresholding technique is disclosed when the background scene is mixed with different clutter sub-groups. Some of these features may have means larger than the target mean, while some of them may have means smaller than the target mean. This technique selects a lower bound threshold (below the target mean) and a higher bound threshold (above the target mean). This technique in combination with the spatio-temporal fusion techniques will threshold out most of the different clutter groups. Further, a reverse-thresholding technique is disclosed for use when the background scene contains non-stationary clutters with increasing or decreasing means. The detection assignment criteria may be reversed depending on if the clutter mean is larger or smaller than the target mean.

## DESCRIPTION OF THE DRAWINGS

[0009] Fig. 1 is a block diagram illustrating the relationship between temporal fusion and spatial fusion;

[0010] Figs. 2a-2d are graphs depicting the performance of pre-detection and post-detection integration;

[0011] Fig. 3 is a block diagram illustrating target detection involving two-targets;

[0012] Fig. 4 is a graph depicting the Gaussian probability of detection of a target and clutter noise;

[0013] Figs. 5a-5b are graphs depicting the performance of single frame detection for receiver operating characteristics;

[0014] Figs. 6a-6d are graphs depicting additive spatio-temporal fusion;

[0015]   Figs. 7a –7d are graphs depicting additive spatio-temporal fusion;

[0016] Figs. 8a –8d are graphs depicting additive and MIN fusions;

[0017] Figs. 9a –9d are graphs depicting a persistency test;

[0018] Figs. 10a –10d are graphs depicting an additive fusion and persistency test;

[0019] Figs. 11a –11d are graphs depicting a combination  additive fusion and persistency

test;

[0020] Figs.  12a –12e  are  graphs  depicting  auto-correlations  of  real  and  computer

generated noise;

[0021] Figs. 13a –13d are graphs depicting noise de-trend;

[0022] Figs. 14a –14d are graphs depicting target detection using real IR sensor noise;

[0023] Figs. 15 a and 15b are graphs depicting the combination of pre-detection and post-

detection with real IR sensor noise for single target and two target cases;

[0024] Fig. 16 is a graph depicting Gaussian pdfs of clutters and targets;

[0025] Fig 17 is a block diagram of a hardware system that performs spatiotemporal

fusion;

[0026] Fig 18 is a flow chart  for implementing temporal fusion utilizing pre-detection

integration;

[0027] Fig 19 is a flow chart for implementing temporal fusion utilizing post-detection

integration; and

[0028] Fig 20 is a flow chart for implementing temporal integration and spatial fusion

from an IR sensor and an RF sensor.

## DETAILED DESCRIPTION OF THE INVENTION

[0029] There are a number of acronyms associated with the description of the present invention, and in order to facilitate an understanding of the description, a glossary of acronyms is provided below:

ATR - automatic target recognition

CFAR - constant-false-alarm-ratio

FPA - focal plane array

FPN - fixed pattern noise

IR – infrared

NUC - non-uniformity correction

Pd - probability of detection

Pdf - probability density function

Pfa - probability of false-alarm

ROC - receiver operating characteristics

RV - random variable

STD – standard deviation

SCNR - signal-to-clutter-noise-ratio

[0030] Although techniques of the present invention are aimed for improving target detection, these techniques can be used for other applications involving thresholding techniques. In target recognition, ATR (automatic target recognition) is a research area with high attention. One popular ATR approach uses the matched filtering/correlation techniques, and the resulting features after the correlation (e.g., the peak-to-sidelobe-ratio) will subject a threshold-screening to pick the recognized targets. Therefore, both the pre-

and post-detection temporal integration methods can be used to enhance target recognition when multiple temporal frames are involved.

[0031]   The assignee of the present invention has a number of currently pending patent applications related to the subject matter of the present invention.   These pending applications include patent application SN 10/395,215, filed March 25, 2003, entitled "Method and System for Multi-Sensor Data Fusion Using a Modified Dempster-Shafer Theory", by Chen et al.; patent application SN 10/395,264, filed March 25, 2003, entitled "Method and System for Target Detection Using an Infra-Red Sensor", by Chen et al.; patent application SN 10/395,265, filed March 25, 2003, entitled "Method and System for Multi-Sensor Data Fusion ", by Chen et al.; patent application SN 10/395,269 filed March 25, 2003, entitled "Method and System for Data Fusion Using Spatial and Temporal Diversity Between Sensors", by Chen et al.;   all of which are incorporated herein by reference.

[0032] The  present invention involves sensor clutter noise looking at real scenes, such as trees, grass, roads, and buildings, etc. Typically, the sensor clutter noise at most of the sensor pixels in a scene, usually more than 95% of the pixels, is near stationary. The sensor clutter noise is un-correlated between pixels, as well as almost being un-correlated across time frames. The noise at a few pixels has shown non-stationary properties with an increasing or decreasing mean across time.   Pixels with these non-stationary properties could include pixels that represent the grass near the edge of a road.

[0033]  If clutters with broader pdf (probability density function) than the target are encountered, it is desirable to determine whether the broad clutter pdf is caused by non-stationary noise with a time-variant mean or is caused by a mix of different clutter types with different stationary means.  Then different detection techniques, such as the double-thresholding or reverse-thresholding schemes, may be selected accordingly.

[0034] Temporal correlation and non-stationary properties of sensor noise have been investigated using sequences of imagery collected by an IR (256x256) sensor looking at different scenes (trees, grass, roads, buildings, etc.). The natural noise extracted from the IR sensor, as well as noise generated by a computer with Gaussian and Rayleigh distributions have been used to test and compare different temporal integration strategies. The simulation results show that both the pre- and post-detection temporal integrations can considerably enhance target detection by integrating only 3~5 time frames (tested by real sensor noise as well as computer generated noise). Moreover, the detection results can be further enhanced by combining both the pre- and post-detection temporal integrations.

[0035] For a physical sensor, the sensing errors are mainly caused by the measurement noise $n_m$ that is generally described as a random variable (RV).  For example, for an IR (infrared) sensor, the measurement noise (temporal noise) may originate from a number of sources including the scene background, atmosphere transmission, path radiance, optics, filters, sensor housing and shield, detector dark current, pixel phasing, quantization, amplifier and read-out electronics, etc.

[0036] For target detection at the feature level, different features are extracted from the original physical measurements. In the IR sensor, for detecting a resolved target occupying multiple pixels of for an unsolved target occupying only a single pixel, a spatial matched filtering process in general is conducted before the detection (thresholding) process. The filter can be a Sobel edge extractor, a difference of Gaussian filter, a specific tuned basis function, or an optical point spread function. The output of the filter is considered the feature values for detection.

[0037] The extracted features affected by the measurement noise are also RVs. The pdf (probability density function) of a feature RV may or may not have the same distribution as the original measurement noise. If a measurement noise has a Gaussian distribution and the extracted feature is a linear transform (e.g., the mean or average of multiple data points is a linear feature) of the physical measurement, the distribution of the feature RV will still be Gaussian. On the other hand, if the relationship between the extracted feature and the original measurement is non-linear, the feature distribution, in general, will be different from the original one. For example, for a radar sensor with a Gaussian distributed measurement noise, if we use the amplitude of the radar return real and imaginary signals as the extracted feature, the distribution of the feature RV will be Rayleigh. To increase the Pd (probability of detection), we must reduce the influence of the feature RVs. The influence of RVs can be decreased by reducing the variances ($\sigma^2$) of the RVs and/or by increasing the distance ($d$) between the means of the two feature RVs related to the target and the clutter). The reduced feature variances and/or the increased feature distances will increase the signal-to-clutter-noise-ratio (SCNR) and thus lead to a better ROC (receiver

operating characteristics) performance, i.e., a higher Pd for the same Pfa (probability of false alarms).

[0038]    Two approaches for reducing the variance of RVs are 1) temporal integration between time frames by averaging the RVs in different frames (the pre-detection integration), and 2) a binomial persistency test using a window of time frames (the post-detection integration). Wold in 1938 proposed and proved a theorem. See, Haykin, Simon, "Adaptive Filter Theory, Prentice-Hall Inc. 1986. This theorem gives us some insight into how temporal integration can be useful:

*Wold's Fundamental Theorem:*

*Any stationary discrete-time stochastic process {x(n)} may be expressed in the form*

$x(n) = u(n) + s(n),$

*where u(n) and s(n) are uncorrelated process, u(n) is a RV, and s(n) is a deterministic process.*

[0039] Therefore, if u(n) is less temporally correlated, temporal integration will be more useful to reduce the variance of u(n). In this case, temporal integration across multiple time frames (temporal fusion) can enhance detection and classification results. The integrated spatio-temporal fusion, which is sketched in Fig. 1, includes a first set of sensors 101 in which there is temporal fusion between frames. There can also be spatial fusion between the first set of sensors 101 and a second set of sensors 102..

[0040] Besides the temporal uncorrelated noise condition that is important for effective temporal integration (fusion), there is another condition need to be addressed. In many realistic situations, the target may be moving and the sensor platform may be moving relative to the background clutters. Therefore, another critical condition for effective temporal fusion is the accurate tracking and associating the targets and clutter objects (i.e., the detected objects) at different time frames using navigation initial tracker and/or image-based tracker or any effective image/object registration/association/correlation techniques.

[0041] We will now describe four fusion (RV combination) strategies: 1) additive, 2) multiplicative, 3) minimum ("MIN"), and 4) maximum ("MAX") fusion.   A more detailed description of the additive fusion and its advantage when adaptively weighting different sensors is provided in Chen et al., "Integrated Spatio-Temporal Multiple Sensor Fusion System Design," SPIE Aerosense, Proceedings of Sensor and Data Fusion Conference, vol. 4731, pp. 204-215, April 2002;  Chen et al., " Adaptive Spatio-Temporal Multiple Sensor Fusion, ", Journal of Optical Engineering, Vol. 42, No. 5, May 2003.

**Additive Fusion**

[0042] The additive fusion rule for two sensors (or two time frames) is

$$p(t) = p(t1) + p(t2), \quad and \quad p(c) = p(c1) + p(c2), \tag{1}$$

where $p(t)$ is the fused target feature values, $p(t1)$ and $p(t2)$ are the target feature values at sensor1 and sensor2 (or time frame1 and frame2), respectively; $p(c)$ is the fused clutter

feature values, $p(c1)$ and $p(c2)$ are the clutter feature values at sensor1 and sensor2 (or time frame1 and frame2), respectively. In a frame, there are generally many more clutter feature values at different pixel locations.

[0043] The additive fusion can be easily extended to include more than two sensors (spatial fusion) or more than two time frames (temporal integration):

$$p(t) = p(t1) + p(t2) + \cdots + p(tn), \quad and \quad p(c) = p(c1) + p(c2) + \cdots + p(cn).$$

(2)

For two independent RVs: X and Y, the combined pdf of the summation of these two RVs $(Z = X + Y)$ is calculated as the convolution of the two individual pdfs:

$$f_Z(z) = \int_0^\infty f_X(x) f_Y(z-x) dx. \tag{3}$$

[0044] In our additive fusion case (with two sensors or two frames), $p(t) = z$, $p(t1) = x$, and $p(t2) = y$ [or $p(c) = z$, $p(c1) = x$, and $p(c2) = y$]. From Eq. (3), we have

$$f_{p(t)}(p(t)) = \int_0^\infty f_{p(t1)}(p(t1)) f_{p(t2)}(p(t) - p(t1)) dp(t1), \tag{4}$$

and

$$f_{p(c)}(p(c)) = \int_0^\infty f_{p(c1)}(p(c1))f_{p(c2)}(p(c) - p(c1))dp(c1).$$  (5)

Eqs. (4) and (5) can be used to predict the detection performance of the additive fusion, since the ROC curves after the additive fusion can be estimated from the combined pdfs in Eqs. (4) and (5).

**Multiplication Fusion**

[0045] The multiplicative fusion rule of two sensors (or two time frames) is

$$p(t) = p(t1) * p(t2), \quad and \quad p(c) = p(c1) * p(c2).$$  (6)

For two independent RVs: X and Y, the combined pdf of the multiplication of these two RVs ($Z = X * Y$) is calculated as the nonlinear convolution (with divisions of a RV) of the two individual pdfs:

$$f_Z(z) = \int_0^\infty \frac{1}{|x|} f_X(x) f_Y(\frac{z}{x})dx.$$  (7)

In our two-sensor multiplication fusion case, from Eq. (7), we have

$$f_{p(t)}(p(t)) = \int_0^\infty \frac{1}{|p(t1)|} f_{p(t1)}(p(t1))f_{p(t2)}(\frac{p(t)}{p(t1)})dp(t1),$$  (8)

and

$$f_{p(c)}(p(c)) = \int_0^\infty \frac{1}{|p(c1)|} f_{p(c1)}(p(c1)) f_{p(c2)}(\frac{p(c)}{p(c1)}) dp(c1). \qquad (9)$$

**The Relationship Between Additive and Multiplication Fusions**

[0046] If we take the logarithm on both sides of the multiplication fusion equations [Eq. (6)], we have

$$ln[p(t)] = ln[p(t1)] + ln[p(t2)], \quad and \quad ln[p(c)] = ln[p(c1)] + ln[p(c2)].$$

(10)

[0047] The multiplication term becomes two additive terms of logarithm functions in each of the equation. If we have two RVs with log-normal pdfs, the equations above indicate that the multiplicative fusion of two RVs with log-normal distributions is equivalent to the additive fusion of two RVs with normal distributions.

## MIN and MAX Fusions

[0048] The conjunction (AND) and disjunction (OR) are two frequently used combination rules in Fuzzy Logic. For two independent RVs: X and Y, the combined pdf of the conjunction of these two RVs [Z = min(X, Y)] is given as

$$f_Z(z) = f_X(z)[1 - F_Y(z)] + f_Y(z)[1 - F_X(z)], \tag{11}$$

where $F(z)$ is the cumulative distribution function.

[0049] Similarly, for two independent RVs: X and Y, the combined pdf of the disjunction of these two RVs [Z = max(X, Y)] is given as

$$f_Z(z) = f_X(z)F_Y(z) + f_Y(z)F_X(z). \tag{12}$$

For our two-object problem, the MIN (conjunction) fusion is

$$p(t) = min[p(t1), p(t2)], \quad and \quad p(c) = min[p(c1), p(c2)]. \tag{13}$$

The MAX (disjunction) fusion is

$$p(t) = max[p(t1), p(t2)], \quad and \quad p(c) = max[p(c1), p(c2)]. \tag{14}$$

[0050] The terms of pre-detections and post-detection integrations were originally used in radar sensor detection. They can be equally applied for IR sensor detection. For both

methods, a temporal moving integration widow (typically containing several frames, e.g., N = 5 or 7) is first selected. In the pre-detection method, one of the different fusion strategies discussed above is applied for the frames within the window size. The fused features values are then used for detection (applying thresholding). In the post-detection (also called persistency test) method, detection (thresholding) is first performed on each image frame within the moving window (with N frames). Then $k$ $(k \leq N)$ detections are evaluated out of the N frames that occurred for a detected object. For example, for a criteria of 5 out of 7, if an object was detected from 5 or more frames in a moving window with 7 frames, the detected object is considered as a target. Otherwise, it is considered as noise or clutter detection.

[0051] Fig 2(a) shows the pdfs (probability density functions) for a noise and a target in a single frame with STD (standard deviation) = 5. Fig 2(b) shows the pdfs after averaging 25 frames (the pre-detection integration, which is equivalent to the additive fusion). The STDof the pdf's in Fig. 2(b) is reduced by a factor of 5. The accumulated probability curves (the error functions) of the pdfs in Fig 2(a) and (b) are plotted in Fig 2(c), where the solid curves denote the single frame and the dashed curves represent the average of twenty-five frames. For the pre-detection integration, the ROC curves are obtained by directly plotting the accumulated probability curves of the target and noise shown in Fig. 2(c) as the y and x axes, respectively, in Fig. 2(d). For a $k$-out-of-$N$ post-detection integration, the accumulated probability curves need to be transferred to post-detection accumulated probability curves using the binomial equation:

$$P(k:N) = \sum_{j=k}^{N} \binom{N}{k} p^{j}(1-p)^{N-j} \qquad (15)$$

where p is a specific probability value from the accumulated probability curves in Fig. 2

(c). Therefore, all the values of a curve in Fig. 2(c) can be transferred to a new curve use

of equation (15). It can be seen that equation (15) contains all the probabilities of $k$ out of

$N$, ($k+1$) out of $N$, . . . until $N$ our of $N$. A ROC curve for the post-detection integration is

obtained by directly plotting the transferred accumulated probability curves of the target

and the noise as the $y$ and $x$ axes, respectively, in Fig. 2(d).


[0052] Several ROC curves are plotted in Fig. 2(d). The top and bottom (solid and

dashed) curves are 7-frame average and 3-frame average pre-detection integration,

respectively. The middle two (dash-dotted and dotted) curves are 5-out-of-7, and 6-out-

of-7 post-detection integration or persistency test results, respectively. It can be seen from

Fig. 2(d) that for a same frame window (e.g., 7 frame), the pre-detection integration

performs better than the post-detection integration.


[0053] Referring now to Fig. 17, a block diagram illustrates a hardware system that can

be used to implement spatiotemporal fusion of data from a plurality of sensors 10, 11. The

sensors in Fig. 17 include an IR sensor 10 and a RF sensor 11. The sensors do not have be

of different types, and such a system could be implemented using multiple sensors of the

same type. The outputs of the IR sensor 10 and RF sensor 11 are temporally fused using

temporal processors 12, 13, respectively, as described in more detail below. The

temporally fused outputs of temporal processors 12, 13 are then preferably applied to a

spatial processor 14 for spatial fusing and detection. The output of the spatial processor

14 is applied to a utilization device 15. The utilization device 15 could be a simple visual

display or a more complicated device, such as a tracking system or an automatic target recognition system.

[0054] As shown in Fig 3, we simulated both IR and RF sensors for target detection enhancements using spatio-temporal fusion. In Fig. 3, the squares represent target 1, the circles represent target 2 and triangles represent clutter noise. Spatial Fusion (integration) is conducted between the IR and the RF frames (pre-detection integration only), while Temporal Fusion is conducted across several time frames of each sensor (both pre- and post-detection integration). Two target situations were simulated: 1) single-target in the scene, and 2) two-targets in the scene. In general, the single-target case has less adjustable parameters and thus would be easier to compare performances from different fusion strategies than the multiple-target case. However, the multiple-target case occurs in many realistic situations. A two-target case is shown in Fig 3. In this simulation, we used static targets and clutter, and presume perfect object tracking and/or registration across multiple time frames.

[0055] Fifty random data samples (related to fifty time frames) were generated as performance data set for each object (target or clutter noise) to evaluate the detection performance. The detection was conducted using the traditional CFAR (constant-false-alarm-ratio) strategy. For a specific CFAR threshold, each detected target at one of the 50 frames counts on 2% of Pd (probability of detection) for the single-target case, and 1% of Pd for the two-target case. The noise in IR is simulated as a normal distribution with a standard deviation of 10, and the noise in RF is simulated as a Rayleigh distribution with a standard deviation of 6.5. Fig 4 shows the pdfs (probability density functions) of a target

and a clutter noise both with normal distributions. In the single-target case the separation of the means between the target and the clutter noise group is set as $S = m_t - m_c = 19$ and $S = 10$ for rf. In the two target case, $S_1 = 19$ and $S_1 = 25$ or IR; $S_1 = 10$ and $S_1 = 17$ for rf.

[0056] The detection ROC performances without any temporal integration (single frame) are shown in Fig 5 as a baseline performance to compare different temporal fusion strategies. Fig 5(a) shows the baseline result from an IR sensor, while Fig 5(b) shows that from a RF sensor. The y-axis is the Pd (probability of detection), and the x-axis is the false-alarm number per frame. The curve with circle symbols is the result from the single-target case, and the curve with square symbols is the result from the two-target case. It is seen that for a false alarm rate of two false alarms per frame the Pd is about 75% for IR and 87% for RF, and that the single-target case performs a little better that the two-target case.

**Additive Spatial Fusion vs. Additive Temporal Fusion**

[0057] For the four different fusion strategies discussed above, our simulation results for target detection show that the multiplication fusion performs the same as the additive fusion, and the MIN fusion performs better than the MAX fusion. Disclosed herein are the results for the Additive and MIN fusion.

[0058] The detection ROC performance curves for the single-target case of IR sensor are shown in Fig 6(a), while the detection ROC performance curves for the two-target case of IR sensor are shown in Fig 6(b). The curve with the circle symbols shows the baseline performance (single frame). The curve with the triangle symbols shows the result of spatial additive fusion between the IR and the RF sensor, while the curve with the square symbols shows the result of additive temporal fusion by integrating a time window of three frames. Similar results for the RF sensor are shown in Fig 6(c) and 6(d). It is found the spatial fusion improves detection and performs better than the single sensor alone. The IR (the worse sensor) improved more than the RF (the better sensor) did. Furthermore, the temporal fusion using three time frames performs better that the spatial fusion using only two sensors. In general, if the noise in different frames are independent to each other, a temporal fusion with N =2,3,... frames should perform similar to a spatial fusion with N sensors. We will discuss the noise correlation properties between frames below. The results of additive temporal fusion using five time frames are shown in Fig 7. In Fig. 7a 7d, there is a window that is equal to five frames. Fig. 7a depicts the curves for an IR sensor and one target. Fig. 7b depicts the curves for an IR sensor and two targets. Fig. 7c depicts the curves for RF sensor and one target. Fig. 7d depicts the curves for an RF

sensor and two targets.   By increasing the time window of integration, the target detection is further enhanced.

**Additive Temporal Fusion vs. MIN Temporal Fusion**

[0059] The results comparing the additive fusion with the MIN fusion for an integration window of five frames are shown in Fig 8. Both additive and MIN fusions with multiple frames enhance target detection. For the IR sensor (with normal noise distribution), the additive fusion always outperforms the MIN fusion in both the single-target and two-target cases as shown in Fig 8(a) and (b), while for the RF sensor (with Rayleigh noise distribution), the MIN fusion can further enhance target detection, and performs equally well as the additive fusion in both the single-target and two-target cases as shown in Fig 8(c) and (d).

**Post-Detection Integration (Persistency Test)**

[0060] The persistency test has been discussed and shown in Section 4 and Fig 2. Persistency test results for both IR and RF sensors are shown in Fig 9. The three curves in each figure are the persistency test for K out of N frames (K=2,3,4; and N=5). Similar to the result in Fig 2(d), the three curves in Fig 9 show similar detection enhancements.

**Additive Fusion vs. Persistency Test**

[0061] Figure 10 shows the results of additive fusion (the curve with square symbols) and the persistency test (the curve with triangle symbols) for both the IR and RF sensors. It is found from Fig 10 that by integrating only five frames, both additive fusion and persistency test can significantly enhance target detection from the baseline (single frame), with additive fusion performing a little better than the persistency test.

[0062] Furthermore, the additive fusion and the persistency test can be complementary to each other. They can be combined to further enhance target detection. Results using an integration window of five frames are shown in Fig 11. The curves with triangle symbols show the ROC performance of the persistency test, the curves with square symbols show the ROC performance of the additive fusion, and the curves with circle symbols show the combination ROC performance of the additive fusion and persistency test.

[0063] As discussed in the previous sections, the performance of temporal integration depends on the temporal correlation properties of the sensor noise. The better performance can be achieved if the noise across the time frames is less correlated. In the simulate results presented in the previous section, we used computer generated random noise that is generally uncorrelated between frames. What about the real sensor noise? To answer this question, we extracted and studied the multiple frame noise from an *InSb* IR FPA (focal plane array) with 256x256 pixels. Imagery sequences (50 time frames) were collected by this IR sensor looking at different scenes (trees, grass, roads, buildings, etc.).

[0064] Studies of the natural IR noise have revealed that 1) the sensor noise at most (>
95%) of the sensor pixels are near stationary and un-correlated between pixels as well as
(almost) un-correlated across time frames; and 2) the noise at a few pixels (e.g., the grass
aside the road) has shown non-stationary properties (with increasing or decreasing mean
across time). Fig 12(b) shows a typical stationary and uncorrelated noise sequence (50
frames) from a specific pixel. Its auto-correlation function is shown in Fig 12(a). Fig 12(d)
shows a typical non-stationary noise sequence with a decreasing mean across time. Its
auto-correlation function with high temporal correlation is shown in Fig 12(c). Fig 12(e)
shows the auto-correlation function of a Gaussian random noise sequence (50 frames)
generated by a computer (this noise has been used in the simulation discussed in the
previous section). It is seen that the natural noise and the computer-generated noise have
similar auto-correlation functions [Fig 12(a) and (e)], and thus both are highly
uncorrelated across time frames.

[0065] From the natural IR noise, we notice that the non-stationary noise at a specific
pixel always shows high values off the center peak in the correlation function. To
understand whether the high vales caused by the non-stationary properties only, or caused
by both non-stationary and temporal correlation, we have de-trended the non-stationary
noise sequences, and remove the increasing or decreasing means. Then we found that the
de-trended noise (becoming a stationary process) becomes temporally uncorrelated (low
values off the center peak in the correlation function). This finding indicates that the noise
at pixels with high off-center correlation values is non-stationary but not temporal
correlated. One such example of the noise de-trend is shown in Fig 13. Fig 13(a) shows a

non-stationary noise sequence with a increasing mean whose auto-correlation function is shown in Fig 13(b). Fig 13(c) shows the same noise after de-trend process, and its auto-correlation function is shown in Fig 13(d). It is seen that the auto-correlation function in Fig 13(d) has much lower off-center-peak values than that in Fig 13(b). That is, the detrended noise is temporally uncorrelated.

[0066] We have applied the IR real noise to test our different temporal fusion strategies, as well as pre- and post-detection temporal integration. The performances using the stationary IR noise are similar to the performances using computer-generated noise as shown in the previous section. Fig 14(b) shows a stationary target noise sequence (50 frames, the solid curve) and a stationary clutter noise sequence (the dashed curve). The target detection ROC performances are shown in Fig 14(a). The curve with circle symbols shows the baseline (single frame) performance. The curve with triangle symbols shows the performance using persistency test with an integration window of 3 frames (2 out of 3), and the curve with square symbols shows the performance of additive fusion with an integration widow of 3 frames. Fig 14(d) shows a non-stationary target noise sequence (the solid curve) with a decreasing mean and a stationary clutter noise sequence (the dashed curve). The target detection ROC performances are shown in Fig 14(c). It is seen that the detection performances are much worse than the results shown in Fig 14(a).

[0067] The results of combining predetection and postdetection integration with real IR noise for single and two target cases are shown in Figs. 15(a) and 15(b), respectively. The curves with triangles show the ROC performance of the persistency tests with an integration window of three frames, the curves with squares show the ROC performance

of the additive fusion, and the curves with circles show the combined ROC performance of the additive fusion and the persistency test. It can be seen that use of this combination can further improve target detection performance.

**Temporal Fusion and IR Sensor Non-Uniformity Correction**

[0068] In the traditional NUC (non-uniformity correction) design, frame subtraction is generally used to subtract out the FPN (fixed pattern noise). However, direct subtraction of two adjacent frames will double the variance of the temporal noise. To avoid a large increase of temporal noise, the NUC design is applied a feedback loop and only a small fraction of the FPN is subtracted out at each iteration. Nevertheless, if we apply temporal integration in the detection system after the NUC process, we can afford the direct subtraction between two nearby frames, and further reduce the noise. For example, the sum of $n$ original frames results in a variance of $n \times v$ (where $v$ is the single frame variance). On the other hand, the sum of $n$ subtracted frames results in a variance of $2 \times v$, because all the variances in the middle frames are cancelled out and only the two variances in the first and the last frames are leftover. Therefore, for an average of $n$ original frames, the resulting variance is $v / n$, while averaging $n$ subtracted frames, the resulting variance is $2v / n^2$. That is, $(2v / n^2) < (v / n)$ when $n > 2$.

**Double-Thresholding Detection Scheme**

[0069] If the feature values of all different clutters in a scene are lager (or smaller) than the target feature value as indicated in Fig 4, the traditional CFAR detection scheme will still

works. For the example in Fig 4, the CFAR scheme always treats an object with a feature value below the threshold as a clutter, and above the threshold as a target. However, in reality, the clutter situations are very complicated. As shown in Fig 16, some clutter groups (e.g., some trees or roads) may have feature values lower that the target, while some other clutter groups (e.g., some decoy-like objects, or counter-measurement objects) may look more like the target and thus have feature values higher than the target. In these situations, the traditional CFAR scheme will partly fail because it only uses a single-thresholding scheme that can only threshold out one of the clutter groups. This increases the likelihood that other groups will be incorrectly be detected as targets.

[0070] In the situation that some clutter feature values are larger than and some are smaller than the target feature value, we propose a double-thresholding scheme with one up-bound threshold and one lower-bound threshold. The technique in combination with the temporal integration will considerably enhance target detection. For example, as shown in Fig 15, suppose the two clutters and the target have Gaussian distributions with the same variances. The separation of the target from the two clutters is two $\sigma$ (i.e., two standard deviation):

$$m_t - m_{c1} = m_{c2} - m_t = 2\sigma$$

[0071] If we set the double thresholds as one $\sigma$ below and one $\sigma$ above the target mean $m_t$, the detection criteria is that only a object with a feature value larger than the lower bound threshold and smaller than the higher bound threshold is assigned as a detection. This is a two-sigma probability and for a Gaussian distribution the Pd (Probability of target

detection) is around 68%, and the Pfa (probability of false-alarm) caused by the two clutter

groups is around 34% (= 17% + 17%). This is the baseline performance for the traditional

single frame detection. However, if we apply the temporal integration of 9 frames with the

additive fusion (equivalent to averaging 9 frames), the standard deviations for the clutters

and the target will be reduced by a factor of 3. It should be presumed that the noise in the

frames is temporally un-correlated.    Then this is a six-sigma probability. The Pd is

increased to above 99%, and the Pfa caused by the two clutters is reduced to below 2%.


[0072] In this technique, for appropriately selecting the two thresholds, we prefer to have

the pre-knowledge of the target mean that may be available from some good training data.


**Reverse-Thresholding Detection Scheme**


[0073] Another situation that the traditional CFAR scheme will fail is when non-stationary

targets and/or clutters exist. As shown in Fig 14(d) where a non-stationary target with a

decreasing mean exists. At an earlier time moment, the target mean is larger than the

clutter mean, while at a later time moment the target mean is below the clutter mean. For a

traditional CFAR single-thresholding approach, we set a single threshold, and any object

with a feature value above it will be assigned as a detected target.  (It should be noted that

for the traditional CFAR scheme, the threshold itself is changing (floating) from frame to

frame to keep a constant false-alarm rate.)  This approach works at earlier time moments

when the target mean is larger than the clutter mean. However, it will fail when the target

mean moves close to and further below the clutter mean, the clutter will have much higher

probability to be falsely detected as a target than the real target. That is why the detection performances in Fig 14(c) are lesser than those in Fig 14(a).

[0074] Similarly, a non-stationary clutter situation can be easily understood using Fig 15. Suppose at an earlier moment the non-stationary clutter with a increasing mean was at the clutter1 location. At a later time moment, it moved from the left side of the target to the right side of the target at the clutter2 location. Based on these observations, we propose a reverse-thresholding scheme to deal with the non-stationary case. As shown in Fig 15, when the non-stationary clutter mean is blow the target mean, we set the criteria for assigning a detection as the object's feature value is above the threshold, while when the clutter mean changed to above the target mean, we set the criteria for assigning a detection as the object's feature value is below the threshold. This technique needs the real time measurements of the changing mean of a non-stationary process. This task may be conducted by using a temporal moving widow or the Wiener and/or Kalman filtering techniques.

[0075] Referring now to Fig. 18, a flow chart illustrates how temporal fusion utilizing post-detection integration or a persistency test can be implemented. In Fig. 18, the first step 21 is to extract a feature value from the first time frame of a sensor. A threshold technique is implemented in step 22 in order to make a detection from the data output during the time frame. In step 23 it is determined whether a predetermined number of time frames have been processed. If N time frames have been processed, then a determination is made in step 24 whether a certain number of detections have been made.

If the number of detections have been made, then there is a positive indication of detection and the process is ended in step 25..

[0076] Referring now to Fig. 19, a flow chart illustrates how temporal fusion utilizing pre-detection integration can be implemented. In Fig. 19, the first step 31 is to extract a feature value from the first time frame of a sensor. In step 32 it is determined whether a predetermined number of time frames have been processed. If N time frames have been processed, then feature values from the predetermined number of time frames are fused using one or more of the fusion functions described above. A threshold technique is implemented in step 34 in order to make a detection from the data output during the predetermined number of time frames N. If the thresholding technique results in a positive indication of detection, the process is ended in step 35.

[0077] Referring now to Fig. 20, a flow chart illustrates how spatiotemporal fusion utilizing data output from a plurality of sensors can be implemented. In Fig. 20, the plurality of sensors includes an IR sensor and a RF sensor. The first steps 41, 42 include extracting a feature value from the first time frame of each sensor. In steps 43, 44 it is determined whether a predetermined number of time frames have been processed from each sensor. If N time frames have been processed, then in steps 45, 46 feature values from the predetermined number of time frames are temporally fused using one or more of the fusion functions described above. In step 47, the temporally fused data is spatially fused utilizing a fusion function. A threshold technique is implemented in step 48 in order to make a detection from the data generated during the spatial fusion 47. If the

thresholding technique results in a positive indication of detection, the process is ended in step 49.

[0078] The sensor and data fusion techniques described above are effective ways to improve target detection and recognition. Current research in this field concentrates mainly in the direction of spatial fusion (fusion from different sensors). The temporal fusion (i.e., fusion across multiple time frames within a specific sensor) of the present invention can also considerably improve target detection and recognition.

[0079] A parameter for temporal fusion is the fusion window size of multiple time frames. In general, the larger the window size the better the fused results that are achieved. However, under some nonstationary situation or in the presence of large tracking errors (or both), a large window will cause large uncorrelated errors. Both the predetection and postdetection temporal integrations of the present invention considerably improve target detection by preferably integrating only ~3-5 time frames (tested by real sensor noise as well as computer-generated noise). These disclosed predetection temporal integration techniques (additive, multiplicative, or MIN fusion) perform better than the traditional postdetection temporal integration technique (persistency test). Detection results can be further improved by combining both the predetection and postdetection temporal integrations.

[0080] Although most examples disclosed herein are for target detection, the techniques can also be used for target recognition (such as the ATR approach with matched filtering and correlation techniques), provided multiple time frames are available. It should be

noted that fusion is conducted in the feature domain by fusing tracked object features across different time frames, but it is not conducted in the original image domain. For example, if the extracted feature is the peak-to-sidelobe ratio of ATR correlation, the ATR with fused features across multiple time frames will perform better than the ATR with a feature from only a single frame.

[0081] Two advanced thresholding techniques, double thresholding and reverse thresholding, have been disclosed. They should perform well in some complicated clutter situation in which the traditional CFAR single-thresholding technique may fail. A simple example of the double-thresholding technique in a complicated clutter situation with a mix of two clutter types has been disclosed. The double-thresholding technique, in combination with temporal fusion of multiple time frames, can improve the Pd from 68% to 99%. In the actual application of the double-thresholding technique, there should be some prior knowledge of the target mean and distribution to set the upper- and lower-bound thresholds. In general, this information can be obtained from reliable training data. It should be noted, however, that the clutter types may number more than 2 and the noise across the time frames may not be totally temporally uncorrelated.

[0082] The training data suggests that, if clutter groups are encountered with a pdf that is broader than that for the target, then a determination should be made whether the broad clutter pdf is caused by nonstationary noise with a time-variant mean or by a mix of different clutter types with different stationary means. Once this is known, different detection techniques can be selected, such as the disclosed double-thresholding or reverse thresholding schemes.

[0083]   The present specification describes a number of different techniques including temporal fusion, spatial fusing and thresholding  and these techniques can be implemented empirically in various ways and combinations using the principles set forth herein.

[0084] Although the invention is primarily described herein using particular embodiments, it will be appreciated by those skilled in the art that modifications and changes may be made without departing from the spirit and scope of the present invention. As such, the method disclosed herein is not limited to what has been particularly shown and described herein, but rather the scope of the present invention is defined only by the appended claims.